

대역 분할 장구간 예측을 이용한 CELP 부호화기의 성능 향상에 관한 연구

최용수*, 서정태, 윤대희, 차일환
연세대학교 전자공학과

A Study on Improvement of CELP Coder using Band-Divided Long Term Prediction

Yong-Soo Choi*, Jeong-Tae Seo, Dae-Hee Youn, Il-Wan Cha
Dept. of Electronics Eng. Yonsei Univ.

요약

본 논문에서는 4.8 kbps 이하의 낮은 전송률에서 Code-Excited Linear Prediction(CELP) 방법에 Multi-Band Excitation(MBE) 방법을 도입하여 장구간 예측 성능을 향상시키는 방법을 제안하였다.

제안된 방법에서는 기존의 CELP 방법으로 장구간 예측을 한 후에도 여전히 남아 있는 주기적 성분들을 다시 한번 다대역(multiband) 장구간 예측을 한다. 이 때 전체 주파수 영역을 기본 주파수의 하모닉 간격으로 주파수 분할하였고, 주기적 다대역 여기 신호를 각 대역 내의 하모닉들에서 여기 신호 스펙트럼과 근사한 크기를 갖는 정현파(sine wave)의 합으로 표현함으로써 실제 여기 신호의 특성을 잘 반영하도록 하였다.

음질 평가의 비교 기준 대상으로 4.8 kbps DoD CELP를 선정하여 주관적 음질 평가(MOS)를 실시한 결과, 4.8 kbps DoD CELP 보다 우수한 성능을 보임을 알 수 있었다.

I. 서론

이동 통신에 대한 수요가 급증함에 따라 채널을 효율적으로 사용하기 위해서는 낮은 전송률에서 고음질을 유지하는 음성 부호화 알고리즘 연구는 필수적이며, 다른 한편으로 음성 우편(voice mail)과 응답 전화기와 같은 시스템에서도 효율적인 음성 저장을 위해서도 낮은 비트를 음성 부호화에 관한 연구는 매우 중요하다[1].

CELP[2]는 4.8 kbps 이하의 낮은 전송률에서도 비교적 양호한 음질을 얻을 수 있어서 가장 각광 받고 있는 이동 통신용 부호화 방법이다.

일반적인 CELP 알고리즘에서 전송률을 낮추기 위한 손쉬운 방법은 프레임의 길이를 늘리는 것이다. 그러나 이 경우에는 장시간 예측 후에도 피치 성분들이 많이 남아 있게 되어 합성음의 음질은 크게 저하된다. 이러한 문제점들을 해결하기 위하여 여러 방법들이 제안되었다[3]-[5]. 이러한 방법들로부터 합성음의 주기성을 향상시키는 것이 주관적(perceptual) 음질을 높일 수 있다는 사실을 알 수 있다. 이 중에서 [5]에서 제안된 MBCELP(Multiband CELP)는 기존의 CELP처럼 1차(first) 적응 코드북으로부터 장구간 예측을 한 후, 다대역으로 필터링된 임펄스열

로 구성된 2차(second) 적응 코드북으로부터 한번 더 장시간 예측을 하여 잔차 신호에 남아 있는 주기적 성분들을 제거함으로써 합성음의 음질을 크게 향상시켰다. 그러나 MBCELP는 4.8 kbps 이하의 전송률로 구현될 경우에 제한된 비트 수로 인해 다대역의 수가 제한되고, 실제 이상적 여기 신호 스펙트럼은 각 대역 내의 하모닉들에서 크기가 다르게 분포되어 있음에도 불구하고, MBCELP에서의 주기적 다대역 여기 신호 스펙트럼은 각 대역 내의 하모닉들에서 일정한 크기를 갖도록 구성된다.

본 논문에서는 위의 MBCELP와는 다른 접근 방식의 Sinusoidal MBCELP(SMBCELP)를 제안하였다. 제안된 SMBCELP에서는 장구간 예측 후 잔차 신호의 스펙트럼을 기본 주파수의 하모닉 간격으로 대역 분할하고, 주기적 다대역 여기 신호의 스펙트럼을 각 대역 내의 하모닉들에서 실제 여기 신호 스펙트럼과 근사한 크기를 갖는 정현파(sine wave)의 합으로 모델링함으로써 성능 향상을 유도하였다.

본 논문은 다음과 같이 구성된다. 2 장에서는 SMBCELP 알고리즘을 설명하며, 3 장에서는 SMBCELP 모델 파라미터를 추정하는 방법을 보여준다. 4 장에서는 4.8 kbps SMBCELP에 대한 컴퓨터 모의 실험 결과를 고찰하고, 5 장에서 결론을 맺는다.

II. SMBCELP 모델

SMBCELP의 전체적인 구조는 그림 1과 같으며, MBCELP [5]와는 근본적으로 주파수 대역 분할 방법과 주기적 다대역 여기 신호의 생성 방법에 차이가 있다.

단구간 예측기로는 인접 샘플 간의 상관 관계는 제거할 수 있지만 유성음 구간에서의 준주기성(quasi periodicity)을 제거하지는 못한다. 이러한 준주기성을 제거하기 위하여 장구간 예측기를 사용한다. 그러나, 장구간 예측기가 주기적 여기 신호를 나타내는데 커다란 기여를 했음에도 불구하고 예측 후에도 여전히 주기적 성분들이 상당히 남아 있다. 그림 2(b)에서 주어진 장구간 예측 후 여기 신호의 스펙트럼에서 알 수 있듯이 주기적 여기 신호의 스펙트럼 에너지는 주파수 영역에 따라 크기가 다르게 나타난다. 여기서 주목할 것은 주기적 성분들은 그림 2(a)에서 관찰할 수 있는 것과 같이 피치 주기를 갖고 있고, 그 스펙트럼은 그림 2(b)에서 볼 수 있듯이 기본 주파수의 정수배에 해당하는 하모닉들에 에너지가 집중되어 있으며 그 스펙트럼 포락선은 평탄하지 않다는 것이다.

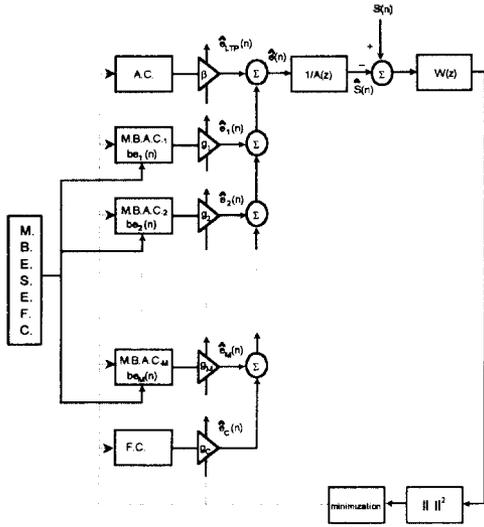


그림 1. SMBCELP 부호화기

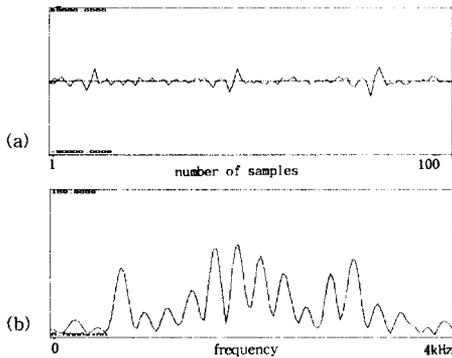


그림 2. 장구간 예측 후의 여기 신호
(a) 신호 파형
(b) 스펙트럼

MBCELP[5]에서는 그림 2(b)에서 보듯이 장구간 예측 후 여기 신호의 스펙트럼이 각 대역 내의 하모닉들에서 크기가 다르게 분포되어 있음에도 불구하고 비트 할당의 제한 때문에 일정한 크기를 갖도록 구성하므로 정확한 모델링이 어렵다.

이를 해결하기 위해 제한된 SMBCELP에서는 주기적 대역역 여기 신호 스펙트럼을 각 대역 내의 하모닉들에서 실제 여기 신호 스펙트럼과 근사한 크기를 갖는 정현파의 합으로 표현한다.

대역역의 최대 개수는 기본주파수의 하모닉들의 개수로 하는 것이 이상적이지만, SMBCELP를 낮은 전송률로 구현하기 위해서 IMBE[6]에서의 유사한 방법으로 대역역들을 다시 몇개씩 묶어서 그 수를 줄일 수 있다.

대역역 적용 코드북 여기 신호 $be_i(n)$ 은 분할된 주파수 대역 B_i 내의 기본 주파수 ω_0 의 하모닉들 ω_0 를 주파수로 하고 크기 A_i 을 갖는 정현파의 합으로 표현될 수 있다.

$$be_i(n) = \sum_j A_j \sin(\omega_0 n + \phi_j(-1)), \quad \omega_0 \in B_i \quad (1)$$

여기서 $\phi_j(-1)$ 은 초기 위상으로 각 하모닉 성분들의 상대적 위상을 보상해주는 역할을 하는데 다음과 같이 표현될 수 있다 [6].

$$\phi_j(0) = \phi_j(-1) + \frac{l(\omega_0(0) + \omega_0(-1))}{2} L \quad (2)$$

여기서 L 과 $\phi_j(-1)$ 과 $\phi_j(0)$ 는 각각 부프레임 길이, 과거 프레임에서의 위상, 현재 프레임에서의 위상이다.

일반적으로 신호 $be_i(n)$ 은 부프레임의 길이 보다 장구간 예측시의 최대 피치 검색 범위 만큼 더 길게 만든다. 첫번째 대역역 적용 코드북에서 최적 위치를 갖는 부프레임 길이의 주기적 여기 신호가 선택된다. 다른 대역역 적용 코드북에 대해서는 여기 신호 벡터의 위치는 첫번째 주기적 여기 벡터의 최적 위치 주위의 작은 범위 내에서 찾아질 수 있거나, 또는 같은 위치로 이득을 다르게 찾을 수 있다.

III. SMBCELP 모델 파라미터의 추정

SMBCELP 모델에서 전송될 파라미터들은 선형 예측 계수, 피치, 장구간 예측(LTP) 이득, 대역역 여기 신호의 스펙트럼 포락선 코드북 인덱스, 대역역 적용 코드북에서의 주기적 여기 신호의 최적 위치, 대역역 주기 여기 신호들의 이득, 통계적 코드북에서 여기 벡터의 최적 인덱스와 이득이다. 단, 두번째 부프레임에서의 피치 파라미터는 첫번째 프레임에서의 피치와의 차이이고, 다른 대역역 적용 코드북에 대한 최적 위치 파라미터는 첫번째 대역역 적용 코드북에서의 주기적 여기 신호의 최적 위치와의 차이이다.

3.1 장구간 예측 파라미터 추정

장구간 예측(LTP) 알고리즘으로 얻어진 장구간 예측 피치는 음성의 실제 피치와 항상 일치하지는 않는다. 그 값은 실제 피치 주기의 정수배 일 수도 있고, 장구간 예측 이득이 음수일 때는 심지어 정수배가 아닐 때도 있다. 따라서 피치가 서서히 변한다 할지라도 인접한 부프레임에서의 장구간 예측 피치의 변화의 차이를 그대로 부호화 하면 음질 저하가 나타난다. 특히, 부프레임 길이가 길어지게 되는 경우에는 합성음의 음질에 더욱 나쁜 영향을 미치게 된다. 이런 영향을 피하기 위해서 매 프레임마다 장구간 예측 피치의 초기값을 추정하는 방법을 사용한다. 원 신호로부터 피치를 추정하는 많은 기법들이 있지만 비교적 신뢰성 있는 IMBE[6]에서의 피치 알고리즘을 사용한다.

각 프레임에서 장구간 예측 피치는 피치 추적에 의해 찾아진 초기 피치를 주위로 아주 작은 범위 내에서 추정된다. 장구간 예측 피치는 천천히 변하므로 두번째 부프레임에 대해서는 장구간 예측 피치 변화의 차이만을 부호화함으로써 비트수를 줄일 수 있다.

3.2 주기적 다대역 여기 신호 패러미터 추정

장구간 예측 후의 여기 신호는 주기성을 갖는 다대역 신호들로 구성되며 장구간 예측 방법과 비슷하다.

다대역 여기 신호 패러미터들은 각 주파수 대역에 대해서 독립적으로 추정될 수 있으며, 분할된 다대역의 갯수에 따라 두가지 접근 방식으로 전개된다.

3.2.1 대역 수가 기본 주파수의 하모닉 갯수와 같을 경우

다대역의 갯수 M 이 기본 주파수 ω_0 의 하모닉들의 갯수와 같을 때는 이상적인 경우로서 MBCELP[5]에서와 같은 패러미터 추정 방법을 사용한다.

3.2.2 대역 수가 기본 주파수의 하모닉 갯수 보다 적을 경우

다대역의 갯수 M 이 기본 주파수 ω_0 의 하모닉 갯수 보다 적을 때는 앞절과는 달리 2장에서 언급한 방법을 사용한다.

각 다대역 내의 기본 주파수의 하모닉에서의 상대적 크기 A_i 는 장구간 예측 후 여기 신호의 스펙트럼 포락선을 기본 주파수의 하모닉 간격으로 샘플링하여 구할 수 있다. 여기서 장구간 예측 후 여기 신호의 스펙트럼 포락선을 추정하는 방법은 다음과 같은 세 가지 방법이 있다.

첫째, 스펙트럼을 기본주파수의 하모닉에서 샘플링하여 피크를 검출하는 방법[7]. 둘째, IMBE[6]에서 유성음 대역에 적용되는 방법인 다대역 스펙트럼과 창합수가 곱해진 임펄스열의 스펙트럼과의 상대적 비로 하는 방법, 셋째, IMBE에서 무성음 대역에 적용되는 대역 스펙트럼의 평균 크기로 하는 방법 등이 있다. 본 논문에서는 이 중에서 주기적 스펙트럼과 잡음 성분이 섞여 있는 스펙트럼 모두에 대해서 스펙트럼 포락선을 비교적 잘 나타낼 수 있는 세번째 방법을 사용하였다.

IV. 실험 및 결과 고찰

4.1 4.8 kbps SMBCELP의 컴퓨터 모의 실험

제안된 방법의 성능 평가를 위해서, 각 패러미터들에 표 1과 같은 비트를 할당하여 4.8 kbps의 전송률을 갖는 SMBCELP 부호화기를 구현하였다.

LPC 분석을 위해 240샘플(30ms)의 Hamming 창합수를 40샘플씩 증첩하여 사용하였다. 한 프레임의 길이는 200샘플(25ms)로 하였으며, 각 프레임은 100샘플(12.5ms)씩 두개의 부프레임으로 구성된다.

4.2 성능 평가

4.8 kbps SMCELP의 성능을 평가하기 위해 비교 기준 부호화기로 4.8 kbps DoD CELP를 선정하였다.

그림 3은 음성 신호의 한 구간에 대한 4.8 kbps SMBCELP 관련 신호들이다. 장구간 예측 후의 잔차 신호 (b)와 다대역 장구간 예측 후의 잔차 신호 (c)를 비교해보면 주기적 성분들이 상당히 제거되었음을 관찰할 수 있다. 그림 4는 그림 3의 스펙트럼이다. 그림 5는 6 초 길이의 한 문장에 대한 4.8 kbps

표 1. 4.8 kbps SMBCELP 패러미터와 비트 할당

Parameters	Bits/frame
LPC(LSP)	34
Pitch1 + Pitch2	7+3
β (LTP gain)	2x4
I_m (MB excit. spec. env. codebook index)	2x6
Pos.1(MB excit.1)	2x7
Pos.2(MB excit.2)	0
g_1 (MB excit. gain1)	2x4
g_2 (MB excit. gain2)	2x4
I_c (codebook index)	2x8
g_c (codebook gain)	2x5
Total	120

SMBCELP 관련 잔차 신호들의 스펙트로그램(spectrogram)이다. 다대역 장구간 예측 후 신호의 스펙트로그램 (b)는 장구간 예측 후 잔차 신호의 스펙트로그램 (a)에 비해서 하모닉 성분들의 에너지가 상당히 줄어들었음을 알 수 있다.

주관적 음질 평가(MOS)에는 21세 전후의 남녀 2명이 각각 두 문장씩 발음한 4개의 음성 시료를 가지고 10명에 대해서 평가 시험을 행하였다. 등급은 1에서 5까지 다섯 등급으로 나누고, 원음을 등급 '5'라고 가정하였다. 그 결과는 표 2에 나타나 있다. 그 결과를 보면 4.8 kbps DoD CELP 보다 약간 높은 점수를 얻었다.

표 2. 주관적 음질 평가(MOS) 결과

	DoD CELP	SMCELP
평균	3.4	3.7

V. 결론

본 논문에서는 4.8 kbps 이하의 낮은 전송률에서 MBE 방법을 도입하여 CELP 부호화기의 장구간 예측 성능을 향상시키는 알고리즘과 패러미터의 추정 방법을 제안하였다.

제안된 방법에서는 기존의 CELP 방법으로 장구간 예측을 한 후, 그 잔차 신호를 에너지가 집중되어 있는 기본 주파수의 하모닉 간격으로 주파수 대역 분할하여 대역에 따라 일정하지 않은 스펙트럼 포락선을 추정하였다. 4.8 kbps 이하의 낮은 전송률로 구현하기 위해 다시 몇 개의 대역들을 묶어서 대역수를 줄였다. 기본 주파수의 하모닉 간격에서 추정된 스펙트럼 포락선을 크기로 하고 기본 주파수의 하모닉을 주파수로 하는 정현파의 합으로 대역 적응 코드북 여기 신호를 만들어 이것으로 대역 장구간 예측을 하였다.

4.8 kbps의 전송률을 갖는 SMBCELP의 실험 결과, 장구간 예측 후의 잔차 신호에 남아 있던 주기적 성분들이 다대역 장구간 예측으로 줄어들었음을 확인하였다. 다대역 장구간 예측 후의 잔차 신호는 하모닉 성분들이 줄어들어서 통계적 코드북으로 여기 신호를 잘 모델링할 수 있다. 그러므로 제안된 방법에서는 CELP 방법에 MBE 방법을 결합하여 실제 여기 신호의 주파수 분포 특성을 잘 반영하였다.

대역 분할 장구간 예측을 이용한 CELP 부호화기의 성능 향상에 관한 연구

제안된 방법의 한 예로서 4.8 kbps SMBCELP를 구현하였으며, 비교 기준 부호화기로 4.8 kbps DoD CELP를 선정하여 음질을 비교 평가하였다. 주관적 음질 평가에서는 4.8 kbps DoD CELP 보다 우수하였다.

주요 과제는 주기적 다대역 여기 신호를 정현파의 합으로 표현할 때, 위상을 더 잘 모델링하여 다대역 장구간 예측 이득을 높이고, 프레임 간 기본 주파수의 변화를 반영할 수 있도록 하는 것이다.

참 고 문 헌

- [1] B.S. Atal, et al., eds., *Speech and Audio Coding for Wireless and Network Application*, Kluwer, 1993.
- [2] M.R. Schroeder, B.S. Atal, "Code-Excited Linear Prediction(CELP) High-Quality Speech at Very Low Bit Rates," *Proc. ICASSP*, pp. 937-940, 1985.
- [3] J.S. Marques and Tribolet, "Pitch prediction with fractional delays in CELP coding," *Proc. EUROSPEECH*, pp. 509-512, 1989.
- [4] S. Miki, K. Mano, H. Ohmuro, and T. Moriya, "A Pitch Synchronous Innovation CELP (PSI-CELP) Coder for 2-4kbit/s," *IEEE ICASSP*, II.113-II.116, 1994.
- [5] G. Yang, H. Leich and R. Boite, "Multiband Code-Excited Linear Prediction (MBCELP) for Speech Coding", *Signal Processing*, Vol. 31, pp. 215-227, 1993.
- [6] IMBE Vocoder description, DVSI, July 1993.
- [7] R.J. McAulay and T.F. Quatieri, "Speech Analysis-Synthesis Based on a Sinusoidal Representation," *IEEE Trans. ASSP*, Vol. ASSP-34, No. 4, August 1986.

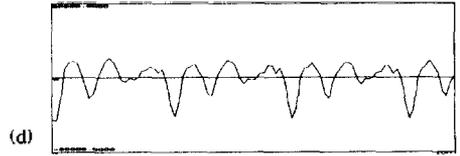
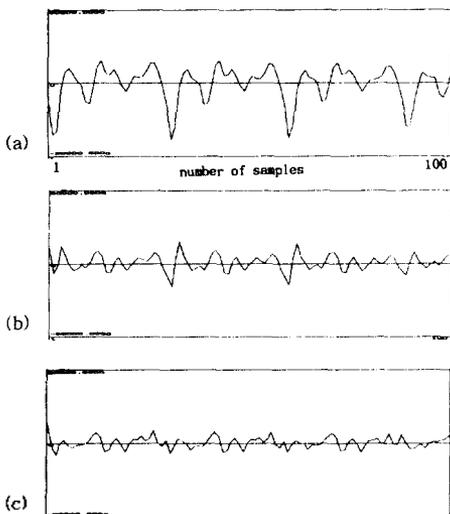


그림 3. 4.8 kbps SMBCELP 관련 신호
 (a) 원 신호
 (b) 장구간 예측 후 잔차 신호
 (c) 다대역 장구간 예측 후 잔차 신호
 (d) 합성 신호

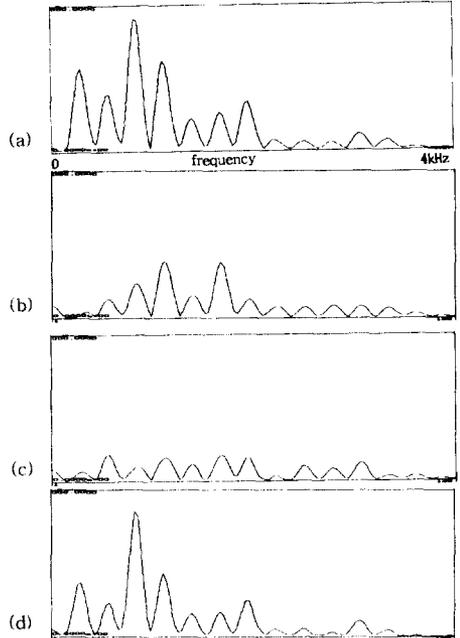


그림 4. 4.8 kbps SMBCELP 관련 신호의 스펙트럼
 (a) 원 신호의 스펙트럼
 (b) 장구간 예측 후 잔차 신호의 스펙트럼
 (c) 다대역 장구간 예측 후 잔차 신호의 스펙트럼
 (d) 합성 신호의 스펙트럼

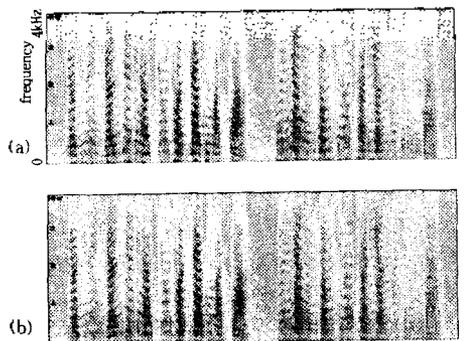


그림 5. 4.8 kbps SMBCELP 잔차 신호의 스펙트로그램
 (a) 장구간 예측 후 잔차 신호의 스펙트로그램
 (b) 다대역 장구간 예측 후 잔차 신호의 스펙트로그램